

Spatial Functional Principal Component Analysis with Applications to Brain Image Data

Yingxing Li *
Chen Huang *²
Wolfgang K. Härdle*³



* Xiamen University, P. R. China

*² University of St. Gallen, Switzerland

*³ Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



Spatial Functional Principal Component Analysis

with Applications to Brain Image Data*

Yingxing Li[†], Chen Huang[‡], Wolfgang K. Härdle[§]

Abstract

This paper considers a fast and effective algorithm for conducting functional principle component analysis with multivariate factors. Compared with the univariate case, our approach could be more powerful in revealing spatial connections or extracting important features in images. To facilitate fast computation, we connect Singular Value Decomposition with penalized smoothing and avoid estimating a huge dimensional covariance operator. Under regularity assumptions, the results indicate that we may enjoy the optimal convergence rate by employing the smoothness assumption inherent to functional objects. We apply our method on the analysis of brain image data. Our extracted factors provide excellent recovery of the risk related regions of interests in human brain and the estimated loadings are very informative in revealing the individual risk attitude.

Keywords: Principal Component Analysis; Penalized Smoothing; Asymptotics; functional Magnetic Resonance Imaging (fMRI).

*Financial support from the Deutsche Forschungsgemeinschaft via IRTG 1792 "High Dimensional Non Stationary Time Series", Humboldt-Universität zu Berlin, is gratefully acknowledged. R codes are accessible via <http://quantlet.de>. More details about the platform can be found in Borke and Härdle (2016, 2017).

[†]The Wang Yanan Institute for Studies in Economics, Xiamen University, China.

[‡]Faculty of Mathematics and Statistics, University of St. Gallen, Switzerland.

[§]Ladislav von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Germany. Sim Kee Boon Institute for Financial Economics, Singapore Management University, Singapore.

1 Introduction

Principal component analysis (PCA) and its functional version (FPCA) are widely used for dimension reduction. This method has been successfully applied in many fields including genetic studies, risk management, psychology, environmental studies, etc (Jolliffe, 2002; Ramsay and Silverman, 2005). The basic principle is to find a basis for a k -dimensional affine linear subspace that best approximates the data. If the data points are finite-dimensional vectors, the basis vectors are called principal components, or factors. If the data points are in an infinite-dimensional Hilbert space, the basis functions are then called functional principal components or factors. One could summarize the data variation and information via some low dimensional loadings, which are projections of the individual observations over the factors.

In classical FPCA, factors are often used to capture the dynamics in time and are assumed as smooth univariate functions of some time variable, say t (Yao et al., 2003; Di et al., 2009; Staicu et al., 2014). Along with the advance of technology, it is now very often to obtain data sets that are measured at different locations or units. It is then of great research interest to characterize the spatial connection among these observations for better inference or prediction. From the perspective of dimension reduction, it is quite natural to extend the idea of FPCA and let the factors be the space function to capture the spatial correlation. However, location index is usually of multi-dimensionality and require careful handling to balance between flexibility and complexity.

One motivated example for this is from neuro-economics study. Recently, there have been explosive interests in understanding which part of our brain is activated during risky decisions and how the human brain regulates specific decision-making tasks (Wang et al., 2013; Majer et al., 2016). A potential solution to this problem is to analyze the fMRI image data during an experiment, where each fMRI image contains signals that are captured

on a 3 dimensional spatial brain maps. In the study of van Bömmel et al. (2014), each fMRI image has a total dimension of $91 \times 109 \times 91$ observations. Such huge dimensionality raises new challenges for statistical inference. A straightforward approach would vectorize the 3D image into 1D vector of dimension 10^6 , but this might implicitly break the spatial connections (Heller et al., 2006). Chen et al. (2015) proposes to conduct a 3D FPCA, but their approach needs to estimate a 6D smooth covariance function, and the computation cost is quite expensive.

In this paper, we propose a general algorithm that could conduct FPCA for d -way arrays, where the spatial feature is captured by nonparametric factors, and individual characteristics is summarized by the factor loadings. Our contributions could be summarized as the follows: First, our approach could be viewed as a nonparametric multidimensional PCA. On one hand, we avoid vectorization and potential destroy of spatial structure. On the other, the nonparametric approach frees us from severe model misspecification. Second, by utilizing penalized splines and singular value decomposition techniques, we propose a fast data-driven algorithm that could reduce computation burdens without much scarification of estimation efficiency; Third, we also develop asymptotic properties for our estimators. To the best of our knowledge, this might be the first paper that discuss theoretical properties of fPCA associated with non 1D domain. Our settings embrace scenarios with a very large dimensional observations and a relatively small number of individuals. Under regularity assumptions, the results indicate that we may enjoy optimal convergence rate by employing smoothness assumption. The real data analysis also demonstrate the success of our approach.

The rest of the paper is organized in the following way. We introduce our estimation procedure and discuss its theoretical properties in Section 2. Section 3 conducts real data analysis and Section 4 concludes. All proofs and technical details are put in the appendix.

2 Methodology

Let $\{X_i\}_{i=1}^I$ denote i.i.d. random functions on a compact region \mathcal{F} , satisfying $\int_{\mathcal{F}} \mathbb{E}(X_i^2) < \infty$. In this paper, we focus on the dense setting such that X_i has d -dimensional array structure, where the total dimension for each individual $n = \prod_{k=1}^d n_k$ could be large. For example, X_i could be the 3D image of individual i , then it consists of an $n_1 \times n_2 \times n_3$ array data. Denote the mean function of all X_i as μ . Then we may represent each X_i as

$$X_i(\mathbf{s}) = \mu(\mathbf{s}) + \sum_{j=1}^{\infty} \psi_{ij} \phi_j(\mathbf{s}), \quad (2.1)$$

where \mathbf{s} denotes the d -dimensional coordinate, ϕ_j and ψ_{ij} are the j th factor and its loading respectively, satisfying that ψ_{ij} are i.i.d. random variables with mean 0 and variance σ_j^2 . In practice, one has noisy observations $Y_{i,\mathbf{s}} = X_i(\mathbf{s}) + \varepsilon_{i,\mathbf{s}}$ instead, where the error terms are independent with 0 mean and variance $\sigma_i^2(\mathbf{s})$.

It is natural to estimate the factors $\phi_j(\mathbf{s})$'s via eigendecomposition performed on a smooth estimate of $\text{Cov}(X_i(\mathbf{s}), X_i(\mathbf{s}'))$. Utilizing multivariate smoothing techniques, we could achieve this task by smoothing the sample version of the covariance $\text{Cov}(Y_{i,\mathbf{s}}, Y_{i,\mathbf{s}'})$ for $\mathbf{s} \neq \mathbf{s}'$. Such an approach could guarantee the orthogonality among $\phi_j(\mathbf{s})$, and it is also valid under irregular sparse setting such that each individual only has a bounded number of observations. However, in the scenario of a regular dense setting where the observations from each individual are of a huge dimension n , both multivariate smoothing and eigendecomposition face great computational challenges.

Therefore, it calls for an algorithm that could circumvent the difficulty mentioned above. As an initial try, we first consider the univariate case where the collection of all $Y_{i,\mathbf{s}}$ forms the vector $Y_i = (Y_{i1}, \dots, Y_{in})^\top$. Under the regular dense setting, the sample covariance has the matrix form $I^{-1} \sum_{i=1}^I (Y_i - \bar{Y})(Y_i - \bar{Y})^\top$, where $\bar{Y} = I^{-1} \sum_{i=1}^I Y_i$. Motivated by the

idea of Xiao et al. (2013), a fast bivariate smoothing could be done via left and right multiplying the smoothing matrix S_λ using a common smoother λ due to the symmetry of the covariance function, where λ is the smoothing parameter and

$$S_\lambda = B(B^\top B + \lambda P)^{-1} B^\top, \quad (2.2)$$

with B and P being the B-spline design matrix and penalty matrix respectively. Correspondingly, the smoothed covariance estimate has the form $S_\lambda \{I^{-1} \sum_{i=1}^I (Y_i - \bar{Y})(Y_i - \bar{Y})^\top\} S_\lambda = I^{-1} \sum_{i=1}^I \{S_\lambda(Y_i - \bar{Y})\} \{S_\lambda(Y_i - \bar{Y})\}^\top$. This implies that, when I is finite or moderately small, the sample covariance of the smoothed individual functions could also yield a smoothed covariance function. In the general case with d way array data $Y_{i,\mathbf{s}}$, we could consider the following algorithm to avoid $2d$ dimensional multivariate covariance smoothing and eigendecomposition:

- Step 1 Partition the whole domain into $m = \prod_{k=1}^d m_k$ cubes and calculate the sample average $\mu_{i,\tilde{\mathbf{s}}}$ and the sample variance $\sigma_{i,\tilde{\mathbf{s}}}^2$ for each cube of index $\tilde{\mathbf{s}}$.
- Step 2 Calculate the $I \times I$ inner product matrix V , whose (i, i') th element is $v_{ij} = m^{-1} \langle \tilde{Y}_i - \tilde{Y}, \tilde{Y}_j - \tilde{Y} \rangle$ for $i \neq i'$, where \tilde{Y}_i is the vector that collects all $\mu_{i,\tilde{\mathbf{s}}}$, and $\tilde{Y} = I^{-1} \sum_{i=1}^I \tilde{Y}_i$. Let V_σ be the diagonal matrix whose (i, i) th element is $\int \sigma_{i,\tilde{\mathbf{s}}}^2 d\tilde{\mathbf{s}}$. Perform eigendecomposition on $\tilde{V} = V - mn^{-1}V_\sigma$ and obtain the raw estimate of the loading $\tilde{\psi}_{ij}$'s.
- Step 3 Regress $Y_{i,\mathbf{s}}$ on loading $\tilde{\psi}_{ij}$ over all i to obtain the raw estimate of the j th factor $\tilde{\phi}_j(\mathbf{s})$. Then smooth $\tilde{\phi}_j(\mathbf{s})$ over \mathbf{s} to obtain $\hat{\phi}_j(\mathbf{s})$.
- Step 4 Regress $Y_{i,\mathbf{s}}$ over $\hat{\phi}_j(\mathbf{s})$ to obtain the updated loadings $\hat{\psi}_{ij}$'s.

In brief, the above algorithm gains computation advantages by combining the idea of singular value decomposition (SVD) and smoothing. Step 1

conducts simple binning to smooth all the observations using a common but raw smoother, while step 2 implements SVD in order to reduce the computation burden caused by conducting eigendecomposition. However, a direct application of SVD on the original data or the smoothed data might still be less effective due to the existence of measurement errors. Therefore, we calculate \tilde{V} and use it for correction. Smoothing in step 3 is crucial for satisfactory estimates of $\hat{\phi}_j(\mathbf{s})$'s. To reduce computation burden, we recommend using the penalized spline techniques and shall provide more detailed discussions later. Step 4 updates the estimated loadings to improve efficiency. Our procedure shares some similarity with Kneip and Utikal (2001) in step 1 and 2, but their work is based on kernel approach, and is not motivated by the connection between smoothing the bivariate covariance and smoothing the univariate individual function. The main distinction is in step 3, where they propose to first smooth the original data (using some common smoother) and then obtain the estimated factors by regression. In contrast, we propose to first regress (using the original data) and then conduct smoothing. We could, indirectly, view our estimated factors as constructed from regressing some pre-smooth data on the loadings, but the amount of pre-smoothing varies across individuals as it implicitly takes into account the impact of the loadings. Moreover, the kernel procedure might be computationally expensive and Kneip and Utikal (2001) use a plug in bandwidth that requires estimates of higher order derivatives. Their approach is valid for the univariate case, but it casts some doubts how it could be extended for multivariate case.

Now we provide more discussions on how our procedure above could be implemented in general for array structure data of arbitrary fixed dimension. Recall that the binning in step 1 aims at getting a crude but consistent estimate of ψ_{ij} , so the requirement on the bin number m is relatively mild. In contrast, the multivariate smoothing in step 3 is essential as it completes the refinements. Hence we first discuss how to implement step 3 using a fast

data-driven algorithm. In general, denote the observation at location \mathbf{s} as $Z_{\mathbf{s}}$, where the coordinate \mathbf{s} contains d components s_1, \dots, s_d , and $1 \leq s_i \leq n_i$. Let Z be the array that collects all $n = \prod_{i=1}^d n_i$ observations. Extending the idea of Xiao et al. (2013), we tempt to utilize the advantages of the tensor product for reducing the computation costs and smooth the covariates such that the fitted data satisfying

$$\text{vec}(\widehat{Z}) = (S_{\lambda_d} \otimes \cdots \otimes S_{\lambda_1}) \text{vec}(Z), \quad (2.3)$$

where S_{λ_j} is the smoothing matrix defined as in equation (2.2) with the penalty parameter λ_j , the B-spline design matrix B_j and the penalty matrix P_j . Different from the bivariate case where one could simply obtain $\widehat{Z} = S_{\lambda_1} Z S_{\lambda_2}$ due to the property of kronecker product, the computation for the general dimension d might be more involved as no sandwich form is available now. Xiao et al. (2013) recommend to compute this by a sequence of nested operations with rotations of array as described in the Generalized Linear Array Model (GLAM) algorithm (Currie et al., 2006). Now we want to propose a new approach to gain more computation advantages by avoiding the rotated transformation for dimension flattening and reinstating.

To illustrate our idea, let us first take a deeper look at the bivariate smoothing. For a given j , notice that the j th column of $S_{\lambda_1} Z$ could be interpreted as fitted values obtained from smoothing $Z_{s_1, j}$ for all $1 \leq s_1 \leq n_1$. Therefore, the left and right multiplying of S_{λ_1} and S_{λ_2} are essentially conducting two conditional smoothing procedures along the first and the second axis respectively. In the general case of d way array, when the smoothing parameter $\lambda = (\lambda_1, \dots, \lambda_d)$ is given, we could also conduct multivariate smoothing via a sequence of d times conditional smoothing, where each time, we condition on all other axis except the i th one, and smooth the available n_i observations using S_{λ_i} . Interestingly, the fitting is irrelevant to the order of conditioning. For example, in the trivariate smoothing case, we could either

conduct three univariate smoothing along each axis in turn, or we could first condition on the third axis and perform bivariate smoothing, and then condition on the first two axes and perform univariate smoothing on the third axis.

To select the appropriate smoothing parameter λ , we propose to use the generalized cross validation (GCV) criterion with the following fast computation expression.

Proposition 2.1. *Let Z be an array structure data of dimension $n = \prod_{i=1}^d n_i$. Suppose the p_i th degree B-splines defined on K_i knots are used. Denote the spline design matrix and the penalty matrix for the i th axis as the matrix B_i and P_i respectively. Then the GCV value equals*

$$\frac{\langle Z - \widehat{Z}, Z - \widehat{Z} \rangle}{n - \text{tr}(S_{\lambda_d} \otimes \cdots \otimes S_{\lambda_1})} = \frac{\langle W \circ W, \widetilde{Z} \circ \widetilde{Z} \rangle - 2\langle W, \widetilde{Z} \circ \widetilde{Z} \rangle + \langle Z, Z \rangle}{n - \prod_{j=1}^d \text{tr}_j}, \quad (2.4)$$

where $\text{tr}_j = \sum_{i=1}^{K_j+p_j} 1/(1 + \lambda_j \eta_{j,i})$ with $\eta_{j,i}$ be defined under equation (A.1), and W and \widetilde{Z} are defined as in equations (A.2) and (A.3).

Remark 2.1. *According to Proposition 2.1, we could precalculate $\eta_{j,i}$'s, \widetilde{Z} and $\langle Z, Z \rangle$. For each λ , we only need to update the calculation of tr_j as well as two inner products that involves W . As the dimension of W is $\prod_{i=1}^d (K_i + p_i)$ instead of $\prod_{i=1}^d n_i$, this could further reduces the computation cost.*

Remark 2.2. *One may also apply the multivariate penalized smoothing along with some data-driven algorithm in step 1. In this case, the pooled GCV criterion proposed by Zhang et al. (2007) might be used, as we adopt a common smoother for all individual functions. For the pooled GCV, it equals a similar formulae as in (2.4) except that we now replace the denominator by $\prod_{i=1}^d (p_i - \text{tr}_i)$.*

Now we consider the asymptotics of our estimates under the following conditions.

- (C1) Suppose $E(Y_{i,\mathbf{s}}^{2+\delta}) < \infty$ and $Y_{i,\mathbf{s}} = X_i(\mathbf{s}) + \varepsilon_{i,\mathbf{s}}$, where the error terms $\varepsilon_{i,\mathbf{s}}$ are independent across individual i and space (\mathbf{s}) . They are also independent of $X_i(\mathbf{s})$ and they have 0 mean and finite variance $\sigma_i^2(\mathbf{s})$. The random function $X_i(\mathbf{s})$ admits representation (2.1) with a bounded support. The factors $\phi_j(\mathbf{s})$ has continuous second derivatives, and the loadings ψ_{ij} are independent with mean 0 and distinct positive variance γ_j for $j = 1, \dots, L$. The number of observations along each axis has the same order $n^{1/d}$, where $n = \prod_{k=1}^d n_k$. The number of individuals satisfies $I = o(n^{d/4})$.
- (C2) Suppose $\tilde{\psi}_{ij}$ are obtained as described in step 1 and 2, where the pilot binning over $Y_{i,\mathbf{s}}$ yields an undersmoothed estimate $\tilde{Y}_{i,\mathbf{s}}$ such that along each axis k , m_k grows at the rate of n_k^δ with $3/4 < \delta < 1$.
- (C3) In step 3, we use a tensor product B -splines basis to estimate $\hat{\phi}_j(\mathbf{s})$. Let $p_k > 0$ and K_k denote the spline degree and the number of knots along the k th axis. The knots are equally spaced on \mathcal{F} and K_k grows faster than the rate of $n_k^{\alpha_k}$ for some fixed constant $\alpha_k > 1/5$. The second order penalty is imposed on $\tilde{\phi}_j$. Denote the penalty parameter as $\lambda_j = (\lambda_{j1}, \dots, \lambda_{jd})$, where the relationship between the penalty parameter and the equivalent bandwidth, along the k th axis, satisfies

$$\lambda_{jk} = \{K_k h_{jk}(In)^{-1/(4+d)}\}^2, \quad (2.5)$$

for some positive constant h_{jk} .

Theorem 2.1. *Assume Conditions (C1)-(C3). For any interior point \mathbf{s} ,*

$$(In)^{2/(4+d)} \{\hat{\phi}_j(\mathbf{s}) - \phi_j(\mathbf{s})\} \xrightarrow{\mathcal{L}} N\{\mathcal{B}_j(\mathbf{s}), \mathcal{V}_j(\mathbf{s})\}, \quad (2.6)$$

where the asymptotic bias $\mathcal{B}_j(\mathbf{s})$ and the asymptotic variance $\mathcal{V}_j(\mathbf{s})$ are given as in equation (A.5).

Remark 2.3. (C1) imposes regularity conditions on the observations. Our approach is similar as a two-stage procedure. (C2) implies that the pilot binning conducted at the first stage has to be so that the bias has negligible effects, while the second stage complements the refinement. The detailed setting of the penalized spline smoothing, including the choice of penalty and the splines are provided in (C3). As pointed out by Li and Ruppert (2008), the placement of the knots is not crucial as long as the number of knots exceeds some threshold. Under this large penalty setting, the key smoothing parameter is the penalty term and the penalized smoothing approach using penalty λ is equivalent to the kernel smoothing approach using an equivalent bandwidth h defined as in equation (2.5).

3 Applications

In this section, we implement our approach to the fMRI data to learn how the human brain responds to the investment decision tasks. In particular, we extract the spatial functional factors to recover the risk related regions of interests (ROIs) and employ the factor loadings to predict the risk attitude on reward and risk.

3.1 Experiment design and data

Our data are collected from a risk perception investment decisions experiment. The data set consists of the high resolution fMRI brain images of 22 subjects (age 18-35 years, 11 females, native German speakers, right-handed and had no history of neurological or psychiatric disease). The same data set is analyzed by Mohr et al. (2010); van Bömmel et al. (2014); Chen et al. (2015). Three participants had to be excluded due to extensive head motion (> 5 mm absolute head movement) or modeling problems (always chose only one alternative). Another two were excluded due to a different scanning fre-

quency. The experiment was conducted as follows: subjects were requested to answer investment related tasks. Each task consists of two phases. In the first phase a stream of 10 returns was sequentially presented for 2×10 seconds. The random return streams were independently drawn from Gaussian distribution with means of 6%, 9%, 12% and standard deviations of 1%, 5%, 9% (9 different combinations in total). In the second phase, subjects performed one of the three task types within 7 seconds: either to choose between 5% fixed safe and risky investment (as shown in the random returns stream) or to tell their subjective judgment (expected return or perceived risk) on the random returns. There were 27 trails for each task type (3×27 tasks in total) and the order of the task types were randomly arranged.

The fMRI data were acquired every 2.5 seconds during the whole experiment procedure. This resulted in 1400 observations of 3-dimensional ($91 \times 109 \times 91$) array that represents the Blood Oxygenation Level Dependent (BOLD) signals. At the same time, the answer for each task from each subject was also collected. The data are available at the Research Data Centre (RDC), Humboldt-Universität zu Berlin.

3.2 Estimation

Recall that for each subject we have a series of 1400 images of 3 dimensional ($91 \times 109 \times 91$). We divide the whole time series into subseries based on the beginning of each task. Since the subject was required to give an answer within 7 seconds and the fMRI was taken every 2.5 seconds, we consider the first three images taken for the l th subject during the m th task, and denote them as $\{Z_{l,m,s,t}\}$, where $1 \leq l \leq 17$, $1 \leq m \leq 81$, $\mathbf{s} = (s_1, s_2, s_3)$ is the spatial coordinate with $1 \leq s_1 \leq 91, 1 \leq s_2 \leq 109, 1 \leq s_3 \leq 91$, and $t \in \{1, 2, 3\}$.

We assume that the subjects exhibit identical brain structure and commonly share the same active regions during the whole experiment procedure.

As we focus on exploring how the BOLD signals change when the subject makes decisions, we take

$$Y_{i,\mathbf{s}} = \frac{Z_{i,\mathbf{s},2} + Z_{i,\mathbf{s},3}}{2} - Z_{i,\mathbf{s},1},$$

where $i = 81 * (l - 1) + m$, or equivalently, $1 \leq i \leq 1377$. The spatial functional factors and the corresponding loadings are estimated according to the algorithm described in Section 2. In step 1, we take the simplest smoothing procedure by evenly dividing the $91 \times 109 \times 91$ voxels into $30 \times 35 \times 30$ cubes and computing the sample mean and variance in each cube, resulting in $\mu_{i,\tilde{\mathbf{s}}}$ and $\sigma_{i,\tilde{\mathbf{s}}}^2$, $\tilde{\mathbf{s}} \in [1, 30] \times [1, 35] \times [1, 30]$, respectively. Then $\mu_{i,\tilde{\mathbf{s}}}$ could serve as approximation to $Y_{i,\mathbf{s}}$ if voxel \mathbf{s} falls into the cube $\tilde{\mathbf{s}}$. Moreover, let V_σ be a diagonal matrix whose (i, i) th element is given by $\int \sigma_{i,\tilde{\mathbf{s}}}^2 d\tilde{\mathbf{s}}$.

Next, we compute an $I \times I$ symmetric matrix V with the (i, i') th element equals $\int \mu_{i,\tilde{\mathbf{s}}} \mu_{i',\tilde{\mathbf{s}}} d\tilde{\mathbf{s}}$. The eigendecomposition $\tilde{V} = V - mn^{-1}V_\sigma = U\Lambda U^\top$ is imposed and the preliminary estimation on the loadings of the j th factor can be obtained by the $\Lambda^{1/2}U_j$, where U_j is the j th column of U . By regressing $Y_{i,\mathbf{s}}$ on $\tilde{\psi}_{ij}$, we obtain the estimate of the j th factors at location \mathbf{s} , i.e. $\tilde{\phi}_j(\mathbf{s})$.

Then we do the 3D penalized smoothing on the factors. We adopt the cubic B -spline bases with 30,35,30 knots along the x, y, z axis respectively. The penalty matrices for all directions are of order 2. Firstly fix s_3 and conduct the 2D smooth on the matrix $\tilde{\phi}_j(\cdot, \cdot, s_3)$. Then fixed (s_1, s_2) and implement the 1D smooth on the vector. After these two steps of conditional smoothing, we achieve $\hat{\phi}_j(\mathbf{s})$. The tuning parameters in the penalized smoothing are selected by minimizing the GCV values calculated from equation (2.4).

At last, we update all the loadings for the i th individual by $\Phi^{-1}G_{\cdot i}$, where the (j, j') th element of the matrix Φ is defined by $\langle \hat{\phi}_j(\mathbf{s}), \hat{\phi}_{j'}(\mathbf{s}) \rangle$ and the j th element of the vector $G_{\cdot i}$ is $\langle \hat{\phi}_j(\mathbf{s}), Y_{i,\mathbf{s}} \rangle$

When implementing the above algorithm, we also need to determine the number of factors and we adopt a two stage procedure for this purpose.

Firstly, we determine a rough choice on the number of factors based on the eigendecomposition of \tilde{V} . Note that \tilde{V} might have some negative eigenvalues due to the numerical approximation in calculation. Let c denotes the magnitude of the minimum eigenvalue. For the eigenvalues with magnitude less than c , we truncate them into 0. For the rest of the eigenvalues, we calculate the cumulative proportion and select the number of factors when the explained proportion first exceeds 80%. After we update the loadings in step S4, we could calculate the variance of the loadings associated with each factor. Then we update our estimate about the number of factors using the same principle, i.e. when the explained proportion first exceeds 80%. In our case, we first select 19 factors and then refine this choice into 9 factors.

3.3 Empirical results

The objectives of our empirical analysis are to recover the risk related regions and also to predict subjects' risk attitude.

3.3.1 Recovery of the risk related ROIs

It is evident from Tobler et al. (2007); Mohr et al. (2010); van Bömmel et al. (2014) that parietal cortex, ventrolateral prefrontal cortex (VLPFC), lateral orbitofrontal cortex (LOFC), anterior insula (aINS), and dorsolateral prefrontal cortex (DLPFC) are the 5 regions which activate the brain activities in investment decision making. They are regarded as the active regions to be recovered in Chen et al. (2015), where the same data set is analyzed. To evaluate the estimation performance of the factors, we also compare the spatial functional factors identified by our approach to those 5 target regions. In particular, the values of the factors are trimmed by the 0.1%- and 99.9%+ quantile levels for each j and the locations where $\hat{\phi}_j(\mathbf{s})$ have nonzero values are marked as red area in Figure 3.1, which presents the recovery of the 5 ROIs.

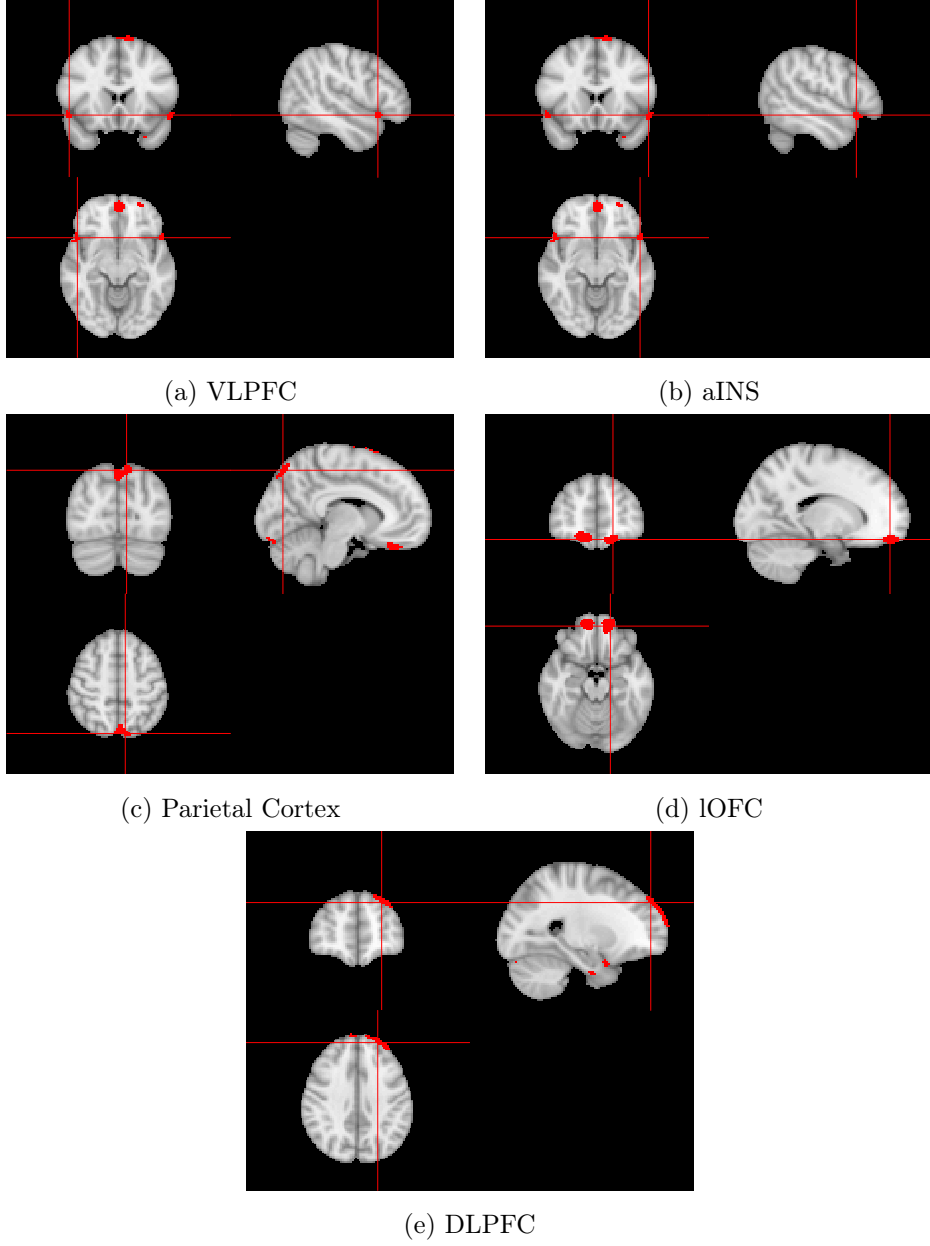


Figure 3.1: The recovery of VLPFC and aINS by $\hat{\phi}_1(\mathbf{s})$, Parietal Cortex by $\hat{\phi}_4(\mathbf{s})$, IOFC by $\hat{\phi}_5(\mathbf{s})$, and DLPFC by $\hat{\phi}_8(\mathbf{s})$, respectively.

The results show that the 5 risk related regions can be identified by $\hat{\phi}_j(\mathbf{s})$ for $j = 1, 4, 5, 8$ with our approach (75.81% of the variance can be explained by the first 8 factors, where the proportion of $j = 1, 4, 5, 8$ is

42.60%). In contrast, Chen et al. (2015) consider the first 19 factors and find that $j = 4, 18, 3(\text{or } 12), 5, 19$ correspond to these 5 regions respectively. Our method seems to be more efficient in extracting the crucial features when dealing with high-dimensional data. In addition, the computational time consuming of our approach is much less compared to Chen et al. (2015), where the exactly same data set is used. In our case, it took 4 hours to run the R codes for our analysis (2.5 hours for loading and preprocessing the data and 1.5 hours for the estimation procedure) on a desktop with Inter(R) Core(TM) i7-6700 CPU @ 3.40GHz processor and 32.0 GB memory without calling parallelization.

3.3.2 Prediction of the risk attitude

We have estimated the subject specific factor loadings and we expect that the individual information carried by the loadings can predict the subjects' risk attitudes. Mohr et al. (2010) estimate the risk attitude parameters R_l for the 17 subjects in the same experiment by the psychological risk-return model, see Appendix B for more details. We shall explore the relationship between their R_l and our estimated loadings by simple regression and conduct predictions on the risk attitudes.

From the brain images associated with the l th subject and the m th task, we obtain the corresponding loading $\hat{\psi}_{i,j}$, $j = 1, \dots, 9$, where $i = 81(l-1) + m$. In order to predict R_l using the loadings, we first summarize the loadings associated with the l th subject and the j th factor by calculating the sample mean and sample standard deviation among the 81 tasks, i.e.

$$\begin{aligned}\nu_{l,j} &\stackrel{\text{def}}{=} \frac{1}{81} \sum_{m=1}^{81} \hat{\psi}_{\{81(l-1)+m\},j}, \\ s_{l,j} &\stackrel{\text{def}}{=} \sqrt{\frac{1}{81} \sum_{m=1}^{81} (\hat{\psi}_{\{81(l-1)+m\},j} - \nu_{l,j})^2}.\end{aligned}\tag{3.1}$$

We shall adopt three approaches to figure a suitable model for risk attitude prediction. To simplify the notation, we drop the subjects indicator l from the subscript to represent the vector accordingly.

In our first approach, we explore the relationship between risk attitude and the covariates of the sample mean ν_j in a standard linear regression model

$$R = \alpha_0 + \sum_{j=1}^9 \alpha_j \nu_j + u. \quad (3.2)$$

As not all covariates might be important, we adopt the backward regression and delete the variable that has the largest p -value respectively until the model selection criterion - AIC no longer decreases. Table 3.1 summarizes the newly deleted variables in each step and the corresponding AIC.

Model	1	2	3	4	5
Deleted variable	-	ν_6	ν_7	intercept	ν_3
AIC	117.74	115.74	113.99	112.60	111.18
Model	6	7	8	9	10
Deleted variable	ν_8	ν_4	ν_2	ν_9	ν_5
AIC	110.90	110.60	108.81	107.51	112.13

Table 3.1: The newly deleted variable and the corresponding AIC of the model in each step by backward regression with ν_j .

The final estimated regression is

$$\widehat{R} = \underset{(0.36)}{-1.50} \nu_1 - \underset{(0.65)}{1.75} \nu_5 + \widehat{u}, \quad (3.3)$$

where the values in the parentheses are the estimated standard deviation errors.

In contrast, if we incorporate the information that factors 1, 4, 5, 8 are associated with the risk related ROIs, then we may start from setting the

regression model as

$$R = \alpha_0 + \sum_{j \in \{1,4,5,8\}} \alpha_j \nu_j + u, \quad (3.4)$$

By backward regression, we also attain the same optimal result as (3.3).

We think that ν_j and ς_j might reflect different information of the overall performance and dispersion among tasks. In our second approach, we plan to include ς_j in the covariates for improvement. We start from the optimal model obtained in the first approach and first use all covariates defined by the sample standard deviation. We then adopt the backward regression and delete the variable that has the largest p -value respectively until the AIC no longer decreases. Table 3.2 summarizes the newly deleted variables in each step and the corresponding AIC.

Model	1	2	3	4	5
Deleted variable	-	intercept	ς_2	ς_3	ς_8
AIC	112.84	110.88	109.17	107.69	106.22
Model	6	7	8		
Deleted variable	ς_6	ς_7	ς_1		
AIC	105.04	103.50	106.36		

Table 3.2: The newly deleted variable and the corresponding AIC of the model in each step by backward regression with incorporating ς_j .

The final estimated regression is

$$\widehat{R} = \underset{(0.48)}{-0.65} \nu_1 - \underset{(1.69)}{2.68} \nu_5 + \underset{(0.70)}{1.34} \varsigma_1 + \underset{(0.89)}{2.60} \varsigma_4 - \underset{(1.58)}{3.4} \varsigma_5 - \underset{(0.89)}{2.14} \varsigma_9 + \widehat{u}. \quad (3.5)$$

In contrast, if we incorporate the information that factors 1, 4, 5, 8 are associated with the risk related ROIs, then we may start from a base model as

$$R = \alpha_0 + \sum_{j \in \{1,4,5,8\}} \alpha_j \nu_j + \sum_{j \in \{1,4,5,8\}} \beta_j \varsigma_j + u. \quad (3.6)$$

By backward regression, we come to the following results

$$\hat{R} = \underset{(0.61)}{-0.98}\nu_1 - \underset{(1.92)}{1.66}\nu_5 + \underset{(0.81)}{1.37}\varsigma_1 + \underset{(1.18)}{2.46}\varsigma_4 - \underset{(1.82)}{3.62}\varsigma_5 - \underset{(1.11)}{1.55}\varsigma_8 + \hat{u}. \quad (3.7)$$

The results are quite similar to our proposed model (3.5), except that it includes ς_8 rather than ς_9 . However, the AIC of this fitted model is 108.00, much higher than the AIC of 103.50 in model (3.5).

In our third approach, we try to further improve our model by adopting a bi-directional stepwise selection based on (3.5). To be more specific, alternately we add the variable which improves the AIC most due to the inclusion of itself and then delete the variable that has the highest p -value (only if the AIC can be further reduced by the exclusion, otherwise we do not take the action). We stop the procedure once the AIC cannot be improved anymore by adding any new variable. All the actions taken in the procedure and the corresponding AIC are presented in Table 3.3.

Model	1	2	3	4
Add(+)/Delete(-)	$+\nu_4$	$+\nu_9$	$+\varsigma_2$	$+\nu_2$
AIC	99.75	99.38	95.62	90.87

Table 3.3: The actions taken in the bi-directional selection and the corresponding AIC of the model in each step.

Note: In each step after a new variable is included into the model, $\varsigma_5, \nu_9, \varsigma_2, \nu_2$ are detected to have the highest p -values accordingly. But the exclusion was not implemented because the AIC was found to become worse.

The final estimation we get is

$$\begin{aligned} \hat{R} = & \underset{(0.55)}{-1.35}\nu_1 + \underset{(1.41)}{0.85}\nu_2 + \underset{(1.57)}{4.07}\nu_4 - \underset{(1.59)}{4.11}\nu_5 + \underset{(2.44)}{4.51}\nu_9 \\ & + \underset{(0.96)}{2.69}\varsigma_1 + \underset{(0.60)}{0.92}\varsigma_2 + \underset{(0.90)}{3.11}\varsigma_4 - \underset{(2.18)}{6.83}\varsigma_5 - \underset{(0.83)}{2.69}\varsigma_9 + \hat{u}. \end{aligned} \quad (3.8)$$

In contrast, if we start from model (3.6) and adopt the bi-direction stepwise

selection, it ends up with the following model

Model	1	2	3
Add(+)/Delete(-)	+ ς_9	- ν_8	+ ν_6
AIC	101.81	99.82	95.53

Table 3.4: The actions taken in the bi-directional selection and the corresponding AIC of the model in each step.

Note: In the last step, ν_6 is detected to have the highest p -values. But the exclusion was not implemented because the AIC was found to become worse.

$$\begin{aligned} \widehat{R} = & \underset{(0.43)}{-0.95} \nu_1 + \underset{(1.50)}{+4.63} \nu_4 - \underset{(1.42)}{3.84} \nu_5 - \underset{(1.22)}{2.31} \nu_6 + \underset{(0.55)}{+1.46} \varsigma_1 \\ & + \underset{(0.84)}{+3.17} \varsigma_4 - \underset{(1.30)}{4.74} \varsigma_5 + \underset{(1.81)}{+4.01} \varsigma_8 - \underset{(1.30)}{5.09} \varsigma_9 + \widehat{u}. \end{aligned} \quad (3.9)$$

For this fitted model, the coefficients associated with $\nu_4, \nu_5, \varsigma_1, \varsigma_4, \varsigma_5, \varsigma_9$ are significant at the level 0.05. Moreover, they are also significant in the estimated model (3.8).

Finally, we compare the out-of-sample performance of the three selection approaches described above. In particular, we conduct a leave-one-out procedure to predict $\{\widetilde{R}_l\}_{l=1}^{17}$. The steps are as below:

- (1) Fix l , where $l = 1, \dots, 17$, conduct ordinary linear regression with the sample of the remaining 16 subjects to estimate the coefficients of the variables in different models.
- (2) Predict \widetilde{R}_l by plugging in the coefficients estimated in step (1).
- (3) Repeat steps (1) and (2) for each $l = 1, \dots, 17$.
- (4) Calculate the Spearman's and Kendall's rank correlations between $\{\widetilde{R}_l\}_{l=1}^{17}$ and $\{R_l\}_{l=1}^{17}$ to check the prediction accuracy on the order of risk-aversion among the subjects.

We summarize the out-of-sample prediction performance of the models in Table 3.5.

Selected variables	Spearman's rank corr	Kendall's rank corr
ν_1, ν_5	0.15	0.13
$\nu_1, \nu_5, \varsigma_1, \varsigma_4, \varsigma_5, \varsigma_9$	0.46	0.35
$\nu_1, \nu_5, \varsigma_1, \varsigma_4, \varsigma_5, \varsigma_8$	0.37	0.26
$\nu_1, \nu_2, \nu_4, \nu_5, \nu_9, \varsigma_1, \varsigma_2, \varsigma_4, \varsigma_5, \varsigma_9$	0.71	0.51
$\nu_1, \nu_4, \nu_5, \nu_6, \varsigma_1, \varsigma_4, \varsigma_5, \varsigma_8, \varsigma_9$	0.57	0.45

Table 3.5: The Spearman's and Kendall's rank correlations for the out-of-sample prediction of the models with the selected variables.

The results in Table 3.5 confirm that the factor loadings estimated in our analysis can provide credible prediction on subjects' risk attitude with appropriately determined models. In particular, we notice that factor 9 might also be of importance for risk perception (especially the dispersion measure of the corresponding loadings), even though it does not match with the five risk related regions; whereas factor 8 might not be as necessary as the other factors associated with the ROIs (factors 1, 4, 5) in predicting subjects' risk attitude.

4 Conclusions

In this paper, we propose the spatial fPCA method that is applicable to arbitrary fixed d way array data. The approach could preserve the spatial structure and efficiently extract the important features via a model free dimension reduction approach. We design a fast algorithm that could reduce the computation burden without sacrificing the estimation efficiency. We also derive the asymptotic properties of our estimator. Potential future research direction could consider impose time dynamic structure on the loadings as well, and thus modelling the temporal and spatial process simultaneously.

APPENDIX

APPENDIX A:

Proof of Proposition 2.1. Recall that the smoothing along the j th axis is $S_j = B_j(B_j^\top B_j + \lambda_j P_j)^{-1} B_j^\top$. If we apply singular value decomposition on B_j^\top and obtain $B_j^\top = U_j D_j Q_j^\top$, then we could define the square root of $(B_j^\top B_j)^{-1}$ simply by $\Sigma_{B_j}^{-1/2} = U_j D_j^{-1} U_j^\top$. Correspondingly,

$$(B_j^\top B_j + \lambda_j P_j)^{-1} = \Sigma_{B_j}^{-1/2} (I_{K_j+p_j} + \lambda_j \Sigma_{B_j}^{-1/2} P_j \Sigma_{B_j}^{-1/2})^{-1} \Sigma_{B_j}^{-1/2},$$

where $I_{K_j+p_j}$ is the identity matrix of dimension $(K_j + p_j) \times (K_j + p_j)$. Perform eigen decomposition to obtain

$$\Sigma_{B_j}^{-1/2} P_j \Sigma_{B_j}^{-1/2} = \tilde{U}_j \tilde{D}_j \tilde{U}_j^\top, \quad (\text{A.1})$$

where \tilde{D}_j is the diagonal matrix whose (i, i) th element is denoted as $\eta_{j,i}$. Then $B_j \Sigma_{B_j}^{-1/2} = Q_j D_j U_j^\top U_j D_j^{-1} U_j^\top = Q_j U_j^\top$, which is an orthogonal matrix. Then the eigendecomposition of S_j could be expressed as

$$\begin{aligned} S_j &= B_j \Sigma_{B_j}^{-1/2} (I_{K_j+p_j} + \lambda_j \tilde{U}_j \tilde{D}_j \tilde{U}_j^\top)^{-1} (B_j \Sigma_{B_j}^{-1/2})^\top \\ &= Q_j U_j^\top (\tilde{U}_j \tilde{U}_j^\top + \lambda_j \tilde{U}_j \tilde{D}_j \tilde{U}_j^\top)^{-1} (Q_j U_j^\top \tilde{U}_j)^\top \\ &=: \hat{U}_j W_j (\hat{U}_j)^\top, \end{aligned}$$

where $\hat{U}_j = Q_j U_j^\top \tilde{U}_j$ is an orthogonal matrix, W_j is the diagonal matrix whose (i, i) th element is $1/(1 + \lambda_j \tilde{D}_j)$.

Let W be an array with dimension $(K_1 + p_1, \dots, K_d + p_d)$ constructed by applying the outer products on the vectors from the diagonal elements of

W_j . For the element of W with coordinate \mathbf{s} being (i_1, \dots, i_d) , we have

$$W_{\mathbf{s}} = \prod_{k=1}^d \left(\frac{1}{1 + \lambda_j \eta_{j, i_k}} \right). \quad (\text{A.2})$$

Let \tilde{Z} be the array obtained from multiplying all \hat{U}_j 's along the j th axis respectively. Equivalently, we could also define \tilde{Z} in an iterative way. When $d = 2$, $\tilde{Z} = \hat{U}_1 Z \hat{U}_2^\top$. When $d = k + 1$,

$$\mathcal{A}(\tilde{Z}) = (\otimes_{j=1}^k \hat{U}_j) \mathcal{A}(Z) \hat{U}_{k+1}, \quad (\text{A.3})$$

where $\mathcal{A}(Z)$ is a mapping that convert the (n_1, \dots, n_d) dimensional array Z into a $(\prod_{i=1}^k n_i) \times n_{k+1}$ matrix such that the k th column collects all observations whose last coordinate equals k . We now show that

$$\langle Z, \hat{Z} \rangle = \langle W, \tilde{Z} \circ \tilde{Z} \rangle. \quad (\text{A.4})$$

First consider $d = 2$. By the definition of \tilde{Z} , we have

$$\hat{Z} = S_1 Z S_2 = \hat{U}_1 W_1 (\hat{U}_1^\top Z \hat{U}_2) W_2 \hat{U}_2^\top = \hat{U}_1 W_1 \tilde{Z} W_2 \hat{U}_2^\top.$$

By the definition of W and the property that W_1 and W_2 are both diagonal matrices, it holds that $W_1 \tilde{Z} W_2 = W \circ \tilde{Z}$. Let $n = \prod_{i=1}^d n_i$. Then

$$\begin{aligned} \langle Z, \hat{Z} \rangle &= \text{vec}(Z^\top) \text{vec}(\hat{Z}) = \text{vec}(Z^\top) \text{vec}(\hat{U}_1 (W \circ \tilde{Z}) \hat{U}_2^\top) \\ &= \text{vec}(Z^\top) (\hat{U}_2 \otimes \hat{U}_1) \text{vec}(W \circ \tilde{Z}) \\ &= \text{vec}(\tilde{Z})^\top \text{vec}(W \circ \tilde{Z}) = \langle \tilde{Z}, W \circ \tilde{Z} \rangle = \langle W, \tilde{Z} \circ \tilde{Z} \rangle. \end{aligned}$$

Assume equation (A.4) holds when $d = k$. Now we prove that it also

holds for $d = k + 1$. By definition of \widehat{Z} and equation (A.3), we have

$$\mathcal{A}(\widehat{Z}) = (\otimes_{j=1}^k \widehat{U}_j)(\otimes_{j=1}^k W_j)\mathcal{A}(\widetilde{Z})W_{k+1}\widehat{U}_{k+1}^\top.$$

As $(\otimes_{j=1}^k W_j)\mathcal{A}(\widetilde{Z})W_{k+1} = \mathcal{A}(W \circ \widetilde{Z})$, we conclude that

$$\langle Z, \widehat{Z} \rangle = \text{vec}(\widetilde{Z})^\top \text{vec}(W \circ \widetilde{Z}) = \langle \widetilde{Z}, W \circ \widetilde{Z} \rangle = \langle W, \widetilde{Z} \circ \widetilde{Z} \rangle,$$

and thus equation (A.4) holds in general.

Using similar techniques, we could also show that

$$\langle \widehat{Z}, \widehat{Z} \rangle = \text{vec}(W \circ \widetilde{Z})^\top \text{vec}(W \circ \widetilde{Z}) = \langle W \circ \widetilde{Z}, W \circ \widetilde{Z} \rangle = \langle W \circ W, \widetilde{Z} \circ \widetilde{Z} \rangle.$$

Finally, the smoother satisfies $\text{tr}(S_{\lambda_d} \otimes \cdots \otimes S_{\lambda_1}) = \prod_{i=1}^d \text{tr}(S_{\lambda_i})$. Moreover, $\text{tr}(S_{\lambda_i}) = \text{tr}(W_i) = \sum_{j=1}^{K_i+p_i} 1/(1 + \lambda_i \eta_{i,j}) = \text{tr}_i$. Therefore, equation (2.4) holds. \square

Proof of Theorem 2.1. Define V_x as the $I \times I$ matrix whose (i, j) th element is $\int X_i(\mathbf{s})X_j(\mathbf{s})d\mathbf{s}$. Let $\widetilde{\gamma}_j$ and γ_j be the j th largest eigenvalue of \widetilde{V} and V_x respectively. By our assumption (C2) on m_k , the bin width along each direction m_k/n_k is $\mathcal{O}(n_k^{-1/4})$. Therefore, we could follow Kneip and Utikal (2001) and show that

$$\|\widetilde{V} - V_x\|_2 = \sup_{\|v\|_2=1} \{v^\top (\widetilde{V} - V_x)^\top (\widetilde{V} - V_x)v\}^{1/2} = \mathcal{O}_p(In^{-1/2}).$$

Hence $\widetilde{\gamma}_j - \gamma_j = \mathcal{O}_p(I^{1/2}n^{-1/2})$, $j = 1, \dots, L$. Define $\psi_j = (\psi_{1j}, \dots, \psi_{Ij})^\top$ and we could decomposed it as γ_j multiplying a unit vector $p_j = (p_{1j}, \dots, p_{Ij})^\top$. Similarly, we could define $\widetilde{\psi}_j$ and decompose it as $\widetilde{\gamma}_j \widetilde{p}_j$. Note that $|\widetilde{p}_{ij} - p_{ij}| = \mathcal{O}_p(I^{-1/2}n^{-1/2})$. Consequently, $\|\widetilde{p}_j - p_j\|_2 = \mathcal{O}_p(n^{-1/2})$ and hence we have $|\widetilde{\psi}_{ij} - \psi_{ij}| = \mathcal{O}_p(n^{-1/2})$, $j = 1, \dots, L$.

Denote the regression coefficient of $Y_{i,\mathbf{s}}$ on ψ_{ij} as $z_{j,\mathbf{s}}$. For each j , we have $z_{j,\mathbf{s}} = \phi_{j,\mathbf{s}} + e_{j,\mathbf{s}}$, where $e_{j,\mathbf{s}}$ are independent across \mathbf{s} with mean 0

and its variance is $I^{-1}\bar{\sigma}_j^2(\mathbf{s})$, where $\bar{\sigma}_j^2(\mathbf{s}) = I^{-1} \sum_{i=1}^I \sigma_i^2(\mathbf{s})$. Denote $h_j = (In)^{-d/(4+d)} \prod_{k=1}^d h_{jk}$. For any interior point $\mathbf{s}_0 = (s_{01}, \dots, s_{0d})$, denote $z_{j,\mathbf{s}}^*$ as

$$z_j^*(\mathbf{s}) = (nh)^{-1} \sum_{s_1=1}^{n_1} \cdots \sum_{s_d=1}^{n_d} z_{j,\mathbf{s}} \prod_{k=1}^d H\left(\frac{s_{0k} - s_k}{h_{jk}(In)^{-1/(4+d)}}\right),$$

where H is the second order kernel function $2^{-1} \exp(-|x|)$. Note that

$$\mathbb{E}[\hat{z}_j(\mathbf{s}) - z_j^*(\mathbf{s})] = \mathcal{O}\left(\max_{1 \leq k \leq d} \{K_k h_{j,k}^{-2} (In)^{-2/(4+d)}\}\right) = o\left(\max_{1 \leq k \leq d} \{h_{j,k}^2 (In)^{-2/(4+d)}\}\right)$$

and $\text{Var}(\hat{z}_j(\mathbf{s}) - z_j^*(\mathbf{s})) = o((Inh)^{-1})$. Since $z_j^*(\mathbf{s})$ is the Nadaraya-Watson estimate with product kernel, we have $(In)^{2/(4+d)} \{\mathbb{E}[\hat{z}_j(\mathbf{s})] - \phi_j(\mathbf{s})\} \rightarrow \mathcal{B}_j(\mathbf{s})$, and $(In)^{4/(4+d)} \text{Var}(\hat{z}_j(\mathbf{s})) \rightarrow \mathcal{V}_j(\mathbf{s})$, where

$$\mathcal{B}_j(\mathbf{s}) = \sum_{k=1}^d h_{j,k}^2 \frac{\partial^2 \phi_j(\mathbf{s})}{\partial^2 s_k}, \quad \mathcal{V}_j(\mathbf{s}) = \left(\prod_{k=1}^d h_{j,k}\right)^{-1} \bar{\sigma}_j(\mathbf{s}) \left\{ \int H^2(u) du \right\}^d. \quad (\text{A.5})$$

Therefore, we could establish the asymptotic normality similar as equation (2.6).

When ψ_{ij} is not available, we will regress $Y_{i,\mathbf{s}}$ on $\tilde{\psi}_{ij}$. Denote the regression coefficient as $\tilde{z}_{j,\mathbf{s}}$. Note that the deviation between $\tilde{\psi}_{ij}$ and ψ_{ij} is of higher order magnitude $\mathcal{O}_p(n^{-1/2})$. When $I = o(n^{d/4})$, such a deviation is negligible compared to $(In)^{-2/(4+d)}$. Therefore, we conclude that smoothing over $\tilde{z}_{j,\mathbf{s}}$ will yield the same asymptotic results as equation (2.6). □

APPENDIX B: Risk Attitude Parameter

The risk attitude parameter R is estimated by logistic model via maxi-

mum likelihood estimation (MLE)

$$\begin{aligned} P\{\text{risky choice}|x\} &= \frac{1}{1 + \exp[-\{\bar{x} - S(x)R - 5\}]}, \\ P\{\text{sure choice}|x\} &= 1 - \frac{1}{1 + \exp[-\{\bar{x} - S(x)R - 5\}]}, \end{aligned} \quad (\text{C.1})$$

where x is the displayed return stream, \bar{x} and $S(x)$ are the subjective expected return and perceived risk judged by the subjects.

The estimated risk attitude parameters for the 17 subjects in order are plotted in Figure C.1. Lower parameters imply risk-seeking behaviors; while higher parameters indicate averse risk patterns. The two extremes #19 and #21 are the most risk-averse and most risk-seeking persons respectively. Note that 5 of 22 subjects were excluded from the sample due to the aforementioned reasons, but we keep the numbers remain.

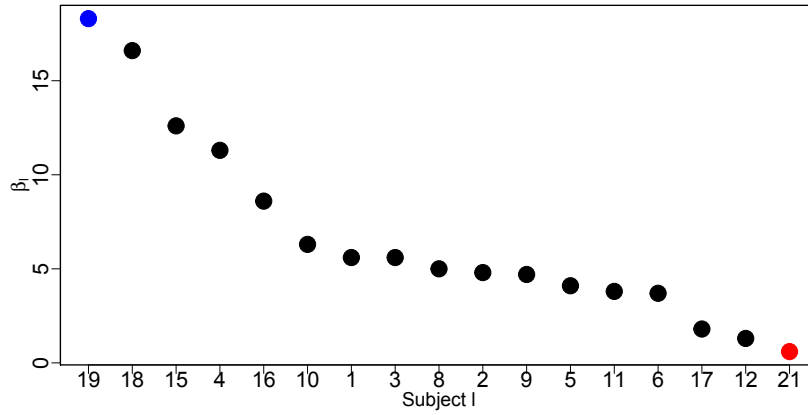


Figure C.1: Risk attitude parameters for the 17 subjects.

References

Borke, L. and Härdle, W. K. (2016). Q3-D3-LSA, *SFB 649 Discussion Paper 2016-049*, Sonderforschungsbereich 649, Humboldt Universität zu Berlin,

- Germany. Available at <https://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2016-049.pdf>.
- Borke, L. and Härdle, W. K. (2017). GitHub API based Quant-Net Mining infrastructure in R, *SFB 649 Discussion Paper 2017-008*, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. Available at <https://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2017-008.pdf>.
- Chen, Y., Härdle, W. K., He, Q. and Majer, P. (2015). Risk related brain regions detected with 3D image FPCA, *SFB 649 Discussion Paper 2015-022*, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. Available at <https://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2015-022.pdf>.
- Currie, I., Durban, M. and Eilers, P. (2006). Generalized linear array models with applications to multidimensional smoothing, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(2): 259–280.
- Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. and Punjabi, N. M. (2009). Multilevel functional principal component analysis, *The Annals of Applied Statistics* **3**(1): 458.
- Heller, R., Stanley, D., Yekutieli, D., Rubin, N. and Benjamini, Y. (2006). Cluster-based analysis of FMRI data, *NeuroImage* **33**(2): 599–608.
- Jolliffe, I. T. (2002). *Principal Component Analysis*, 2nd edn, Springer.
- Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis, *Journal of the American Statistical Association* **96**(454): 519–542.
- Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines, *Biometrika* **95**(2): 415–436.

- Majer, P., Mohr, P. N., Heekeren, H. R. and Härdle, W. K. (2016). Portfolio decisions and brain reactions via the CEAD method, *Psychometrika* **81**(3): 881–903.
- Mohr, P. N., Biele, G. and Heekeren, H. R. (2010). Neural processing of risk, *Journal of Neuroscience* **30**(19): 6613–6619.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, Series in Statistics, 2nd edn, Springer, New York.
- Staicu, A.-M., Li, Y., Crainiceanu, C. M. and Ruppert, D. (2014). Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis, *Scandinavian Journal of Statistics* **41**(4): 932–949.
- Tobler, P. N., O’Doherty, J. P., Dolan, R. J. and Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems, *Journal of neurophysiology* **97**(2): 1621–1632.
- van Bömmel, A., Song, S., Majer, P., Mohr, P. N., Heekeren, H. R. and Härdle, W. K. (2014). Risk patterns and correlated brain activities. multi-dimensional statistical analysis of fMRI data in economic decision making study, *Psychometrika* **79**(3): 489–514.
- Wang, J., Zhu, H., Fan, J., Giovanello, K. and Lin, W. (2013). Multi-scale adaptive smoothing models for the hemodynamic response function in fmri, *The annals of applied statistics* **7**(2): 904.
- Xiao, L., Li, Y. and Ruppert, D. (2013). Fast bivariate p-splines: the sandwich smoother, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**(3): 577–599.
- Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A. and Vogel, J. S. (2003). Shrinkage estimation for func-

tional principal component scores with application to the population kinetics of plasma folate, *Biometrics* **59**(3): 676–685.

Zhang, J.-T., Chen, J. et al. (2007). Statistical inferences for functional data, *The Annals of Statistics* **35**(3): 1052–1079.

SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Fake Alpha" by Marcel Müller, Tobias Rosenberger and Marliese Uhrig-Homburg, January 2017.
- 002 "Estimating location values of agricultural land" by Georg Helbing, Zhiwei Shen, Martin Odening and Matthias Ritter, January 2017.
- 003 "FRM: a Financial Risk Meter based on penalizing tail events occurrence" by Lining Yu, Wolfgang Karl Härdle, Lukas Borke and Thijs Benschop, January 2017.
- 004 "Tail event driven networks of SIFIs" by Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle and Yarema Okhrin, January 2017.
- 005 "Dynamic Valuation of Weather Derivatives under Default Risk" by Wolfgang Karl Härdle and Maria Osipenko, February 2017.
- 006 "RiskAnalytics: an R package for real time processing of Nasdaq and Yahoo finance data and parallelized quantile lasso regression methods" by Lukas Borke, February 2017.
- 007 "Testing Missing at Random using Instrumental Variables" by Christoph Breunig, February 2017.
- 008 "GitHub API based QuantNet Mining infrastructure in R" by Lukas Borke and Wolfgang K. Härdle, February 2017.
- 009 "The Economics of German Unification after Twenty-five Years: Lessons for Korea" by Michael C. Burda and Mark Weder, April 2017.
- 010 "Data Science & Digital Society" by Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, May 2017.
- 011 "The impact of news on US household inflation expectations" by Shih-Kang Chao, Wolfgang Karl Härdle, Jeffrey Sheen, Stefan Trück and Ben Zhe Wang, May 2017.
- 012 "Industry Interdependency Dynamics in a Network Context" by Ya Qian, Wolfgang Karl Härdle and Cathy Yi-Hsuan Chen, May 2017.
- 013 "Adaptive weights clustering of research papers" by Larisa Adamyan, Kirill Efimov, Cathy Yi-Hsuan Chen, Wolfgang K. Härdle, July 2017.
- 014 "Investing with cryptocurrencies - A liquidity constrained investment approach" by Simon Trimborn, Mingyang Li and Wolfgang Karl Härdle, July 2017.
- 015 "(Un)expected Monetary Policy Shocks and Term Premia" by Martin Kliem and Alexander Meyer-Gohde, July 2017.
- 016 "Conditional moment restrictions and the role of density information in estimated structural models" by Andreas Tryphonides, July 2017.
- 017 "Generalized Entropy and Model Uncertainty" by Alexander Meyer-Gohde, August 2017.
- 018 "Social Security Contributions and the Business Cycle" by Anna Almosova, Michael C. Burda and Simon Voigts, August 2017.
- 019 "Racial/Ethnic Differences In Non-Work At Work" by Daniel S. Hamermesh, Katie R. Genadek and Michael C. Burda, August 2017.
- 020 "Pricing Green Financial Products" by Awdesch Melzer, Wolfgang K. Härdle and Brenda López Cabrera, August 2017.
- 021 "The systemic risk of central SIFIs" by Cathy Yi-Hsuan Chen and Sergey Nasekin, August 2017.
- 022 "Das deutsche Arbeitsmarktwunder: Eine Bilanz" by Michael C. Burda and Stefanie Seele, August 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 023 "Penalized Adaptive Method in Forecasting with Large Information Set and Structure Change" by Xinjue Li, Lenka Zbonakova and Wolfgang Karl Härdle, September 2017.
- 024 "Smooth Principal Component Analysis for High Dimensional Data" by Yingxing Li, Wolfgang K. Härdle and Chen Huang, September 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

